W240–W245 *Nucleic Acids Research, 2014, Vol. 42, Web Server issue* *Published online 16 May 2014*
doi: 10.1093/nar/gku394

POSA: a user-driven, interactive multiple protein structure alignment server

Zhanwen Li[1], Padmaja Natarajan[1], Yuzhen Ye[2], Thomas Hrabe[1] and Adam Godzik[1,*]

[1]Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA and [2]School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA

Received January 31, 2014; Revised April 18, 2014; Accepted April 24, 2014

ABSTRACT

POSA (Partial Order Structure Alignment), available at http://posa.godziklab.org, is a server for multiple protein structure alignment introduced in 2005 (Ye,Y. and Godzik,A. (2005) Multiple flexible structure alignment using partial order graphs. *Bioinformatics*, 21, 2362–2369). It is free and open to all users, and there is no login requirement, albeit there is an option to register and store results in individual, password-protected directories. In the updated POSA server described here, we introduce two significant improvements. First is an interface allowing the user to provide additional information by defining segments that anchor the alignment in one or more input structures. This interface allows users to take advantage of their intuition and biological insights to improve the alignment and guide it toward a biologically relevant solution. The second improvement is an interactive visualization with options that allow the user to view all superposed structures in one window (a typical solution for visualizing results of multiple structure alignments) or view them individually in a series of synchronized windows with extensive, user-controlled visualization options. The user can rotate structure(s) in any of the windows and study similarities or differences between structures clearly visible in individual windows.

OVERVIEW

The POSA server performs multiple protein structure alignment using the Partial Order Structure Alignment algorithm described in (1). The algorithm and the original POSA server were designed to study variations of protein structures within diverse protein families and were characterized by two unique features: (i) the ability to compare and classify regions that are conserved only in a subset of input structures and (ii) the ability to account for internal rearrangements (flexibility) in protein structures. The POSA algorithm consists of two main steps: (i) calculating all-by-all pairwise structure alignments between all input structures using the FATCAT algorithm (2) in flexible mode and using the neighbor-joining approach to generate a similarity tree of the input proteins; (ii) using the tree from the first step to guide a series of pairwise alignments of POGs (Partial Order Graphs), representing single structures or groups of aligned structures, until all the input structures are aligned (1).

Multiple protein structure alignment (MPStrA) is an important approach for functional and evolutionary analysis of groups of protein structures. Typically, MPStrA is used to identify the conserved regions that form the common structural core of a protein family. Such alignments are identified by MPStrA algorithms like MUSTA (3), CE-MC (4), MultiProt (5), Matt (6), MISTRAL (7), Smolign (8), 3DCOMB (9), SALIGN (10), msTALI (11), mulPBA (12), DeepAlign (13), MUSTANG (14), PDBeFold (15) and others (16,17). Despite its age, POSA performs reasonably well when benchmarked against recent methods such as msTALI (11) or mulPBA (12), and in addition provides features such as the ability to display ligands and the multiple window display, not available on any other MPStrA servers (see Table 1 for overview of features of several recent MPStrA web servers and the supplemental material for detailed comparison of the results for the zinc finger example used throughout this manuscript). Results of various MPStrA algorithms often differ from each other because different algorithms use different scoring systems to evaluate structural similarity and use different heuristics to solve a computationally complex (NP-Complete) problem (18). In addition, with increasing evolutionary distances, structures become more divergent and thus more difficult to align. As a result, the common core found by POSA as well as by other existing MPStrA algorithms could be smaller than suggested even by a quick visual analysis, and sometimes programs fail to find any common core altogether. This problem is exacerbated by the fact that all existing programs typically run in a fully automated mode, with a user being able to control only general parameters, such as gap penalty, thus ignoring the insights an experienced user could have about the structure conservation in a given family. We believe that such insights

*To whom correspondence should be addressed. Tel: +1 858 646 3168; Fax: +1 858 795 5249; Email: adam@godziklab.org

© The Author(s) 2014. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

are very valuable, and rather then being ignored, they could be used to guide the multiple protein structure alignment to a biologically relevant solution.

Here, we describe a major update of the POSA server that we believe achieves this goal by introducing two significant improvements to the original server: (1) a novel, interactive interface that allows the user to guide the algorithm toward the optimal alignment using his or her knowledge and insights about the structures and (2) versatile visualization options that allow the user to view all the aligned structures in one window, which is a typical solution for visualizing results of multiple structure alignments, or view them individually in a series of synchronized windows. In addition, the new interface allows the user to compare the interactions between the aligned structures and their respective ligands or other protein chains in the complex by optionally including them in the visualization. In this manuscript we describe the new features, provide a narrated example of the server's application, and discuss some potential uses of the new features.

SERVER WORKFLOW

Sample dataset

In the following paragraphs we will use a sample dataset of six structures, each with two-tandem C2H2-type zinc finger domains (PDB IDs: 2drp, 2d9h, 2adr, 4gzn, 1x6e and 2cot) as an example to illustrate the POSA server interface and its features. A detailed Help section available on the server provides additional examples and describes other possible workflows.

Input

The main entry to the POSA server is an input interface in which the user enters the PDB IDs and chain designation of input proteins and other optional program parameters (Figure 1A). Coordinates can also be uploaded from a file. The remaining columns in the interface are optional, providing access to the special features of the POSA server. In particular, the 'Segments' column is used to define one or more segments, which will limit the POSA alignment to the desired section of the input structures. The 'Use' feature will be illustrated on the example of the zinc finger dataset used throughout this manuscript. The simplified multiple structure alignment can be calculated by the progressive alignment method (19) to a selected specific structure by enabling the 'Reference' field. In addition, the user can select 'Other Chains' for each structure. These chains will not be used in the alignment but will be included in the output visualization, maintaining their relative positions to their respective 'master' chains. In our example, displaying the DNA chain along with the aligned zinc finger chains allows one to compare the zinc finger—DNA binding between different zinc finger structures. Six input rows are available by default but can be extended to more rows matching the user's needs until a limit of 20 chains is reached. Multiple alignments between a larger number of input structures are possible but would exceed the visualization capabilities of the current server. Finally, the 'Show more input methods' box can be used to access additional options for POSA input, which are documented on the server Help page.

The zinc finger example. The input interface comes prefilled with sample input values for the zinc finger dataset used as an example here (Figure 1A). It provides an option for a simple, single-click submission of a sample job to demonstrate the typical workflow of the server. In this example, rigid structure alignment of the input structures in the automated mode, with no segments and reference specified, fails to find a common core. A common core is detected, however, when segments are specified for each structure: segment definitions used here focus the alignment on the first zinc finger domain in all the structures. Furthermore, we specified chains B, C and A, B from the protein structures 2drp and 4gzn, respectively, to display protein–DNA interactions in the resulting alignment.

Output

After submitting a job, the POSA server will display a page reporting the job status until the job is completed. A unique, private job ID is available for the user to return to the results later. Alignment results are available through links on a job home page (Figure 1B) that also provides a summary of the results. In particular, the job home page provides links to the rigid MPStrA result, information about the common core and the RMSD value for the resulting alignment. An additional link to flexible MPStrA results would be available if structural flexibilities are detected in the input structures. Since no flexibilities have been detected for our zinc finger example, no link to the flexible MPStrA result is available.

Links on the job home page provide access to additional pages with information on Partial Order Alignment (POA), Partial Order Graph (POG), Simplified Partial Order Graph (SPOG), alignment of amino acids and the superposed protein structures in PDB format. All results can be downloaded for local use from the job home page (Figure 1B) and can be accessed from the top menu in the structure visualization page (Figure 1C).

Alignment guide tree. The guide tree displayed on the job home page illustrates the relations between the analyzed proteins (Figure 1B) and could be interpreted as a simple approximation of a true phylogenetic tree. The number of the aligned residues in each branch is indicated on the tree and highlighted in red if flexibility is detected in the alignment. The number of the aligned residues at the root of the tree indicates the length of the common core for all input protein structures. Branches of the guide tree are clickable, and by clicking on a specific branch, the structural alignment for only that branch will be displayed. Thus, the user can inspect the tree to identify branches grouping highly similar protein structures or branches separating group of significantly different protein structures. Users can utilize this information to identify a subset of the input protein structures and reiterate MPStrA.

Alignment POA graph. The POA graph provides a novel visualization of the multiple protein structure alignment. All other types of alignment representations can be derived

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

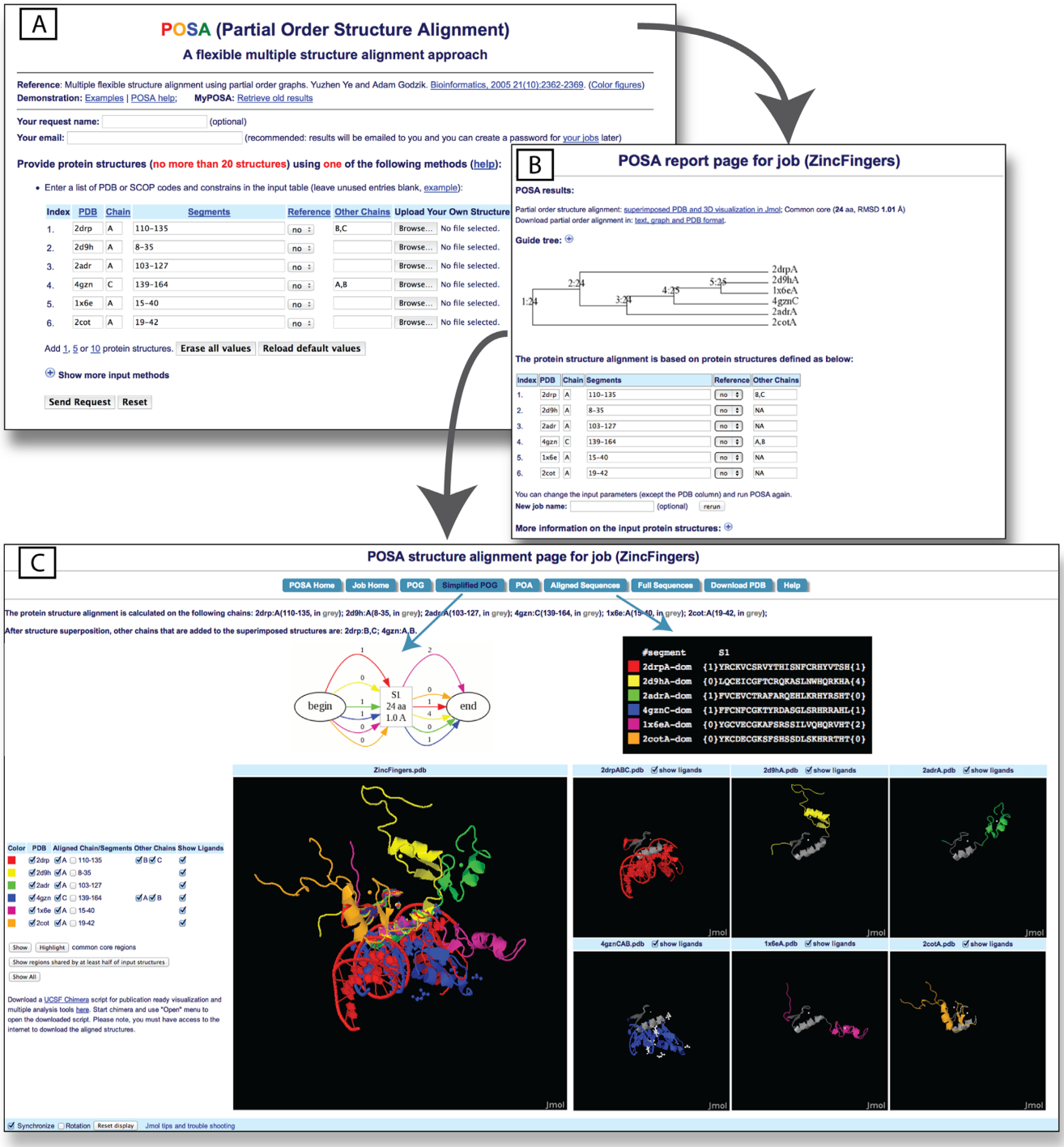

Figure 1. A typical POSA server workflow: (**A**) Input interface to start MPStrA jobs on the POSA server. Users fill in an input table providing PDB IDs of the structures (alternatively uploading coordinates from a file) and the chain IDs. Additionally, the user has an option to define segments in one or more structures to anchor alignment on these specific parts. Specifying 'Other Chains' will not affect the alignment result but will allow these chains to be visualized consistently with the chain or segment(s) being aligned. (**B**) The job home page contains links to the MPStrA result visualization, information about the common core and the RMSD value detected for the alignment. It also displays the alignment guide tree that indicates similarity by branch distance. Clicking on the respective sub-branches will display the alignment of proteins in the chosen sub-branch only. An editable input table with the current settings makes it easy to iteratively refine the current alignment. (**C**) Visualization page of the alignment result. The upper menu provides links to specific result information such as the POG or the sequence of the common core. Below are the Jmol views for visualization of the superposed or individual structures. All views are synchronized by default but can be viewed independently if required. The menu on the left of the main view controls multiple visualization modes. POSA visualization on a large monitor and the arrangement of the Jmol windows may vary for smaller screens.

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

Table 1. A comparison of MStrA web servers features

	POSA	mulPBA	PDBeFold	msTALI	Smolign	SALIGN
Input						
PDB/chain/upload	X/X/X	X/X/X	X/X/X	o/o/X	X/X/X	X/X/X
Specify ligand chains	X	o	X	o	o	o
Guided alignment	X	o	X	o	o	X
Quality scores						
RMSD	X	X	X	X	X	o
Common core length	X	o	X	X	X	o
Online visualization						
Superposed structures	X	X	X	X	X	o
Synchronized views	X	o	o	o	o	o

Only servers that were updated after 2010 were included in the comparison. Features in this table are sorted in categories and X indicates a feature that is available on a given server.

from the POA file, which is a text file in XML format described in detail on the Help pages. The aim of the POA XML is to provide an interface for interested developers to use POSA results for their subsequent analysis. POG is a graphic view of the POA. It shows the common regions in square boxes and variable regions at the graph edges, among any subset of the input structures. The SPOG graph is a simplified version of the POG and displays only the common core regions in all input protein structures (Figure 1B). Amino acid alignment is provided in a block format in which variable regions are described only by a number indicating how many residues are in the region and common regions are described by their amino acid sequences (Figure 1B).

Iterative refinement of the alignment

The POSA server also displays the initial input data on the job home page in a table identical to the one on the input interface on the start page. This feature enables the user to resubmit jobs with some modifications of the input parameters (Figure 1B). For instance, one can specify the amino acid range for a zinc finger domain for one protein structure (PDB ID: 2drp, Chain ID: A) in the 'Segments' column and set the 'Reference' column for this structure to 'yes.' The server will align the defined zinc finger domain in the reference structure to one of the zinc finger domains in other protein structures using the FATCAT method and provide a pseudo-multiple alignment that could be useful for some applications.

Synchronized interactive 3D visualization of multiple structure alignment

Another feature of the updated POSA server is the new visualization page for the superposed multiple protein structures with versatile visualization options (Figure 1C). We will introduce the options on the visualization page using our example of six zinc finger proteins.

The visualization page has two types of frames: (1) the top frame showing protein sequence-related information controlled by the top menu and (2) the bottom frame displaying the 3D viewer of the MPStrA using Jmol (20). The POG and SPOG graphs in the top frame communicate with the aligned structures in the 3D viewer upon user interaction such as clicking on the specific segments of the aligned structures.

The central 3D viewer window shows superposition of all input protein structures and the smaller 3D viewer windows show individual structures. Rotations in all windows are synchronized by default, allowing the user to rotate one structure with other structures simultaneously repeating the same rotation. This option can be disabled if required. The 3D viewer of the aligned structures includes a selection table with clickable boxes that can activate multiple display modes. Figure 2 shows several modes for the alignment of the zinc finger domains. Various visualization combinations of structures, other chains, and ligands are possible. For example, users can analyze two aligned zinc finger structures (2drp, 4gzn) with the DNA chains only (Figure 2C).

On the visualization page we furthermore provide a download link to a UCSF-Chimera (21) script that can recreate the visualization page for generating publication-ready figures or interface with other structure analysis tools.

Summary of the new POSA interface and capabilities

The new POSA interface provides users an opportunity to guide multiple protein structure alignment by using their expert knowledge by specifying protein regions to be aligned, thus focusing MPStrA on these segments. Additional chains or ligands can be specified for any input structure—they are excluded from the alignment process but are displayed with the same transformation matrix as the aligned segments. The visualization page allows the user to manipulate and analyze any subset of superposed structures, along with ligands and other chains from the input structures. A large view with all superposed structures is synchronized with a series of smaller views of all individual structures for user interaction. Thus, with the new interface, POSA can be used to process (i) rigid MPStrA, (ii) flexible MPStrA, (iii) user-guided structure alignment, (iv) inspect ligands, (v)

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

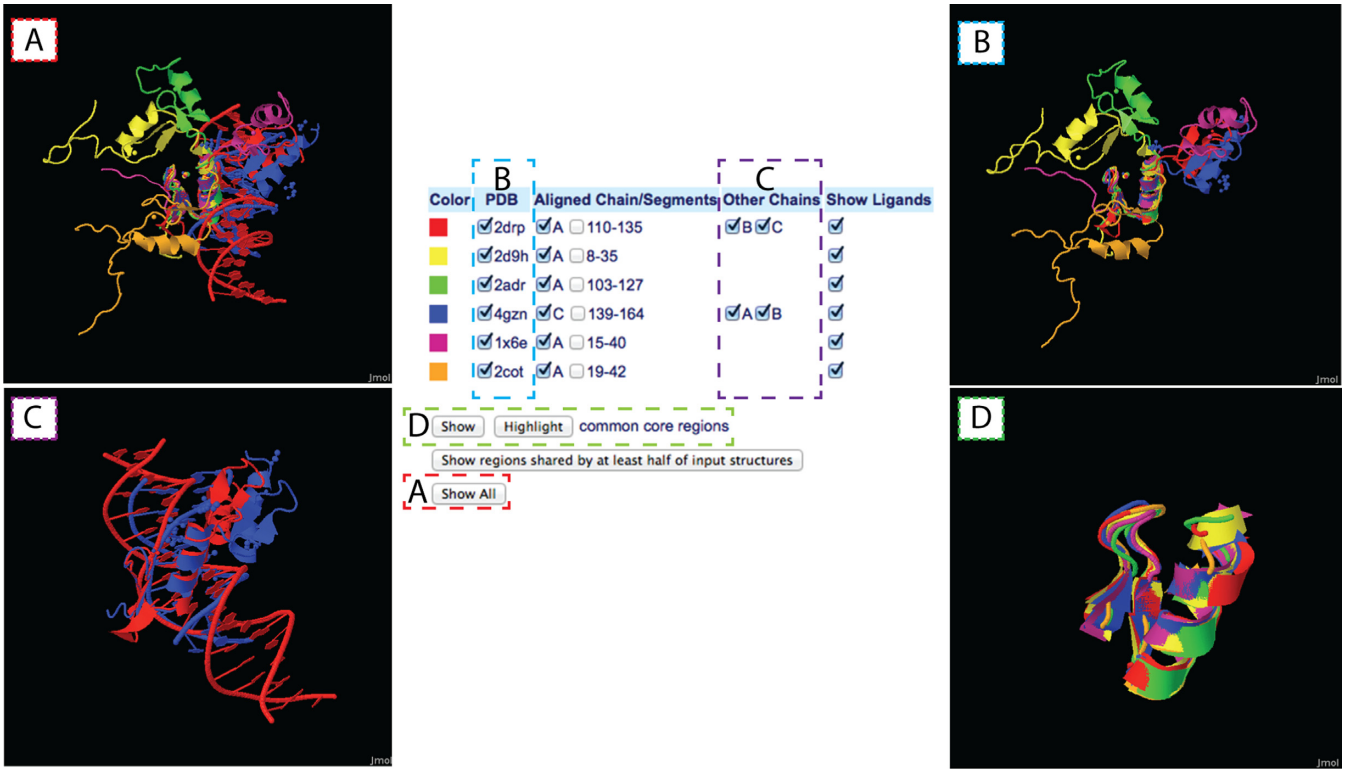

Figure 2. Multiple display modes of the result visualization: (**A**) All zinc finger structures superposed after POSA alignment. All ligands and other chains specified are visible. (**B**) View on the same scene displaying the aligned chains only with all ligands (zinc atoms) visible. (**C**) Only structures for which 'Other Chains' have been specified are displayed. This view mode allows direct comparison of multiple protein–nucleic acid and protein–protein interactions. (**D**) The common core matches the first zinc finger domain of all aligned structures. The highlight button will set all the non-core segments to transparent.

other molecules in the input structures, (vi) analyze protein–nucleic acid and protein–protein interactions.

ACCESSION NUMBERS

PDB IDs: 2drp, 2d9h, 2adr, 4gzn, 1x6e and 2cot.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENT

We thank Drs Alexey Eroshkin and Lukasz Jaroszewski of Sanford-Burnham Medical Research Institute, La Jolla, CA, for their suggestions for the POSA upgrade.

FUNDING

National Institutes of Health [R01GM101457]. Funding for open access charge: National Institutes of Health [R01GM101457] and SBMRI funds.
Conflict of interest statement. None declared.

REFERENCES

1. Ye,Y. and Godzik,A. (2005) Multiple flexible structure alignment using partial order graphs. *Bioinformatics*, **21**, 2362–2369.
2. Ye,Y. and Godzik,A. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, **19**(Suppl. 2), ii246–ii255.
3. Leibowitz,N., Nussinov,R. and Wolfson,H.J. (2001) MUSTA - a general, efficient, automated method for multiple structure alignment and detection of common motifs: application to proteins. *J. Comput. Biol.*, **8**, 93–121.
4. Guda,C., Lu,S., Scheeff,E.D., Bourne,P.E. and Shindyalov,I.N. (2004) CE-MC: a multiple protein structure alignment server. *Nucleic Acids Res.*, **32**, W100–W103.
5. Shatsky,M., Nussinov,R. and Wolfson,H.J. (2002) MultiProt - a multiple protein structural alignment algorithm. *Algorith. Bioinformatics*, **2452**, 235–250.
6. Menke,M., Berger,B. and Cowen,L. (2008) Matt: local flexibility aids protein multiple structure alignment. *PLoS Comput. Biol.*, **4**, e10.
7. Micheletti,C. and Orland,H. (2009) MISTRAL: a tool for energy-based multiple structural alignment of proteins. *Bioinformatics*, **25**, 2663–2669.
8. Sun,H., Sacan,A., Ferhatosmanoglu,H. and Wang,Y. (2012) Smolign: a spatial motifs-based protein multiple structural alignment method. *IEEE/ACM Trans. Comput. Biol. Bioinformatics/IEEE, ACM*, **9**, 249–261.
9. Wang,S., Peng,J. and Xu,J. (2011) Alignment of distantly related protein structures: algorithm, bound and implications to homology modeling. *Bioinformatics*, **27**, 2537–2545.
10. Braberg,H., Webb,B.M., Tjioe,E., Pieper,U., Sali,A. and Madhusudhan,M.S. (2012) SALIGN: a web server for alignment of multiple protein sequences and structures. *Bioinformatics*, **28**, 2072–2073.
11. Shealy,P. and Valafar,H. (2012) Multiple structure alignment with msTALI. *BMC Bioinformatics*, **13**, 105–120.
12. Leonard,S., Joseph,A.P., Srinivasan,N., Gelly,J.C. and de Brevern,A.G. (2014) mulPBA: an efficient multiple protein structure alignment method based on a structural alphabet. *J. Biomol. Struct. Dyn.*, **32**, 661–668.
13. Wang,S., Ma,J., Peng,J. and Xu,J. (2013) Protein structure alignment beyond spatial proximity. *Sci. Rep.*, **3**, 1448.

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015

14. Konagurthu,A.S., Whisstock,J.C., Stuckey,P.J. and Lesk,A.M. (2006) MUSTANG: a multiple structural alignment algorithm. *Proteins*, **64**, 559–574.
15. Krissinel,E. and Henrick,K. (2005) Multiple alignment of protein structures in three dimensionsLecture Notes in Computer Science **3695** 67–78
16. Leibowitz,N., Fligelman,Z.Y., Nussinov,R. and Wolfson,H.J. (1999) Multiple structural alignment and core detection by geometric hashing. In: *Proceedings of International Conference on Intelligent Systems for Molecular Biology*, Vol. **1999**, pp. 169–177. http://www.aaai.org/Papers/ISMB/1999/ISMB99-020.pdf (13 May 2014, date last accessed).
17. Theobald,D.L. and Steindel,P.A. (2012) Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics*, **28**, 1972–1979.
18. Lancia,G. (2008) Mathematical programming in computational biology: an annotated bibliography. *Algorithms*, **1**, 100–129.
19. Feng,D.F. and Doolittle,R.F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.*, **25**, 351–360.
20. Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/. (13 May 2014, date last accessed).
21. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

Downloaded from http://nar.oxfordjournals.org/ at OHIO STATE UNIVERSITY LIBRARIES on December 9, 2015